ELSEVIER

Fisheries Research 58 (2002) 41–57

FISHERIES RESEARCH

www.elsevier.com/locate/fishres

Comparison of logistic and generalized surplus-production models applied to swordfish, *Xiphias gladius*, in the north Atlantic Ocean

Michael H. Prager*

Center for Coastal Fisheries and Habitat Research, National Oceanic and Atmospheric Administration, 101 Pivers Island Road, Beaufort, NC 28516, USA

Received 20 January 2000; received in revised form 19 July 2001; accepted 30 July 2001

Abstract

Recent assessments of swordfish, *Xiphias gladius*, in the north Atlantic Ocean by the International Commission for the Conservation of Atlantic Tunas (ICCAT) have included fitting a nonequilibrium logistic (Schaefer) surplus-production model. The logistic model offers simplicity, but concern has been expressed that its fixed model shape may bias estimates of quantities of management interest. Here, I compare results from the logistic estimator used by ICCAT to those from an otherwise equivalent generalized (Pella–Tomlinson) production-model estimator. Following initial estimation with nonlinear least-squares, a resistant fitting method was used to identify statistical outliers, and both models were refit with outliers removed. The estimate of model shape from the generalized model was then close to the logistic, and estimates of stock status from the two estimators were similar. A simulation study conditioned on the trimmed generalized fit suggests that any systematic estimation error caused by assuming logistic shape for this stock is small. Moreover, the generalized estimator was sensitive to outlying observations and thus less precise than the logistic estimator, and it exhibited larger median proportional unsigned error. Sensitivity to outliers and lack of precision in an estimator make it more likely to provide misleading estimates in a given analysis; therefore, if the generalized production model with estimated shape parameter is used in stock assessment, it should be applied with skepticism and in conjunction with the more robust logistic form. Unless a good external estimate of model shape is available, the logistic model appears more suitable for routine assessment use on stocks similar to swordfish. Published by Elsevier Science B.V.

Keywords: Swordfish; Surplus production; Accuracy; Precision; Robust methods; Outliers; Stock assessment; North Atlantic Ocean

1. Introduction

Choice of model structure in fishery science (as elsewhere) is a process of compromise between simplicity and fidelity to the underlying system (e.g. Punt, 1992). This principle extends even to such simple tools as surplus-production models. In the least complex surplus-production model, the logistic form of Schaefer (1954, 1957), the production curve (curve of production per unit time dB/dt as a function of stock biomass B) is assumed symmetrical around the biomass B_{MSY} that can produce maximum sustainable yield (MSY). In the more complex generalized model (Pella and Tomlinson, 1969), the production curve can be skewed in either direction (Fig. 1). The generalized model is sometimes thought more realistic, or at least more adaptable to possible realities (Pella and Tomlinson, 1969; Quinn

* Tel.: +1-252-728-8760; fax: +1-252-728-8619.
E-mail address: mike.prager@noaa.gov (M.H. Prager).

0165-7836/02/$ – see front matter. Published by Elsevier Science B.V.
PII: S0165-7836(01)00358-7

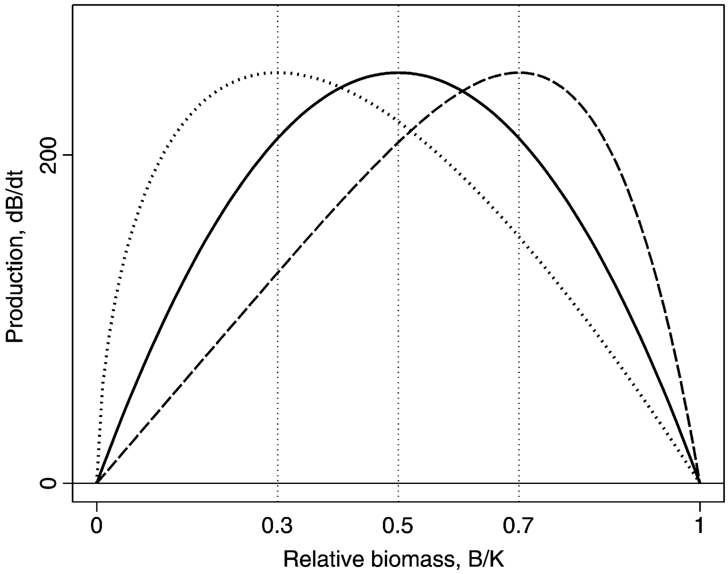

Fig. 1. Typical production model shapes (defined by $\phi \equiv B_{MSY}/K$) for hypothetical populations with MSY = 250. Left curve, $\phi = 0.3$ (equivalent exponent $n = 0.68$); center curve, $\phi = 0.5$ (logistic model, $n = 2$); right curve, $\phi = 0.7$ ($n = 6.04$).

and Deriso, 1999), while the logistic model offers greater simplicity (Lotka, 1924; Prager, 1994).

Since introduction of the generalized model, many authors have noted that estimating its coefficients can be difficult (Pella and Tomlinson, 1969; Fletcher, 1978; Rivard and Bledsoe, 1978). Hilborn and Walters (1992) asserted that "few if any data sets on real fish populations" could support estimation of the shape of the production curve. They further stated: "the Pella and Tomlinson extension of the Schaefer model is a nice theoretical construct but should be used only with great care in fisheries stock assessment". Despite the problems attending parameter estimation, the generalized model continues to attract interest from assessment biologists, an interest that may be sustained in part by availability of the computer program PRODFIT (Fox, 1975), which fits the generalized model by equilibrium approximation. However, the equilibrium approximation and similar methods, although useful when computational power is lacking, have been studied repeatedly and have repeatedly drawn strong criticism (Roff and Fairbairn, 1980; Polacheck and Hilborn, 1987; Hilborn and Walters, 1992; Polacheck et al., 1993).

Management of swordfish *Xiphias gladius* in the north Atlantic Ocean is coordinated by the International Commission for the Conservation of Atlantic Tunas (ICCAT) and conducted by an ICCAT species group, an international panel of scientists representing broad expertise on the species and fisheries throughout its exploited range. In ICCAT's recent assessments of the stock, a nonequilibrium version of the logistic production model, implemented as an observation-error estimator (Pella and Tomlinson, 1969; Hilborn and Walters, 1992; Prager, 1994, 1995), has been used along with other methods, among them age-structured production models (Hilborn, 1990), Bayesian production models (McAllister et al., 2000), and VPA-based catch-at-age models (Gavaris, 1988), sometimes incorporating Monte Carlo simulations to address risk (Restrepo et al., 1992). Although correct shape of the production curve has been a matter of discussion, comparisons of logistic production-model results to those from an equivalent model with estimated shape have not been made, perhaps because of the estimation difficulties mentioned earlier. Some comparison was made to estimates from the Fox (1975) equilibrium-approximation procedure. However, in keeping with recommendations in the literature (Ludwig et al., 1988; Polacheck et al., 1993; Punt and Hilborn, 1996), this ICCAT species group has preferred to avoid equilibrium assumptions and to use observation-error estimators.

In this study, a direct comparison was made between equivalent logistic and generalized estimators, both applied to current data on the stock. Each model was fit through an observation-error estimator (Pella, 1967; Pella and Tomlinson, 1969; Prager, 1994; Punt and Hilborn, 1996) under the same statistical assumptions used by recent ICCAT assessments. The goal was to weigh the simplicity and relative ease of estimation provided by the logistic model against the possibly less biased estimates of the generalized model, and to reach conclusions applicable to the stock of swordfish in the north Atlantic Ocean and to stocks of similar life history and fishing pattern. In the course of the study, the utility of robust regression methods (Rousseeuw and Leroy, 1987) in production modeling was also examined and a simulation was conducted to better evaluate performance of the two estimators in this application.

2. Data and basic methods

Data (Table 1) are the estimates of catch and relative abundance used in ICCAT's 1999 assessment of this stock (ICCAT, 2000). Because catch data include only reported catches, they are considered minimum estimates, and in analysis the proportion of nonreporting must be assumed constant. Relative abundance is estimated by ICCAT through a linear model applied to catch and effort data from Canadian, Japanese, Spanish, and US longline fisheries (Hoey et al., 2000). Data span the years 1950–1998 with several periods of missing relative-abundance data near the start of the series.

2.1. Models

The two models used for analysis were the logistic production model (Lotka, 1924; Schaefer, 1957; Pella, 1967; Schnute, 1977; Prager, 1994) and the generalized production model of Pella and Tomlinson (1969) as restructured by Fletcher (1978). Under the logistic model, a stock's dynamics in the absence of fishing are described by the differential equation:

$$\frac{dB_t}{dt} = rB_t - \frac{r}{K}B_t^2, \quad (1)$$

where B_t is the biomass at time t, r the population's intrinsic rate of increase, and K the limiting population

Table 1
Data used in production models of swordfish in north Atlantic Ocean[a]

Year	Abundance index	Catch
1950	–	3746
1951	–	2781
1952	–	3193
1953	–	3503
1954	–	3134
1955	–	3602
1956	–	3358
1957	–	4578
1958	–	4904
1959	–	6232
1960	–	3828
1961	–	4381
1962	–	5342
1963	1398.86	10,180
1964	501.05	11,258
1965	314.35	8652
1966	293.67	9349
1967	351.31	9107
1968	284.72	9172
1969	258.54	9203
1970	296.39	9495
1971	–	5266
1972	–	4766
1973	–	6074
1974	–	6362
1975	433.51	8839
1976	372.42	6696
1977	392.43	6409
1978	627.49	11,835
1979	355.60	11,937
1980	478.38	13,558
1981	333.95	11,180
1982	385.78	13,215
1983	280.03	14,527
1984	276.91	12,791
1985	266.85	14,383
1986	256.75	18,486
1987	235.11	20,236
1988	232.53	19,513
1989	222.13	17,250
1990	209.34	15,672
1991	217.24	14,937
1992	201.22	15,394
1993	183.19	16,772
1994	166.63	15,235
1995	176.71	16,618
1996	142.68	14,921
1997	147.61	12,913
1998	165.24	12,175

[a] Abundance index in relative biomass units; catch in tonnes.

size (carrying capacity). The units of r are time^{-1} (usually yr^{-1}), while the units of B and K are biomass.

Corresponding dynamics of the generalized model as restructured by Fletcher (1978) are described by the differential equation:

$$\frac{\mathrm{d}B_t}{\mathrm{d}t} = \gamma m \frac{B_t}{K} - \gamma m \left(\frac{B_t}{K}\right)^n, \tag{2}$$

where m is maximum sustainable yield, also symbolized MSY, with units biomass·time^{-1}; n is a unitless exponent determining the shape of the production curve; and γ is a function of n:

$$\gamma = \frac{n^{n/(n-1)}}{n-1}. \tag{3}$$

Equation (2), has a removable singularity at $n = 1$; at that value of n, the generalized model is equivalent to the Fox (1970) exponential yield model.

In the logistic model ($n = 2$), parameters of Eqs. (1) and (2) are related by $m = \frac{1}{4}rK$. Thus Eq. (1) describing the logistic model can be rewritten as:

$$\frac{\mathrm{d}B_t}{\mathrm{d}t} = 4m \frac{B_t}{K} - 4m \left(\frac{B_t}{K}\right)^2. \tag{4}$$

Either model can include fishing by addition of the term $-F_tB_t$ to its right-hand side, where F_t is the instantaneous fishing mortality rate, with units of time^{-1}.

In this study, shape of the production curve is characterized by the unitless ratio $\phi \equiv B_{\mathrm{MSY}}/K$. This seems more biologically natural than using the exponent n itself. The two are related monotonically by

$$\phi = \left(\frac{1}{N}\right)^{1/(n-1)},$$

where n is the exponent in Eq. (2) (Fletcher, 1978). The relationship between the two quantities is markedly nonlinear (Fig. 2).

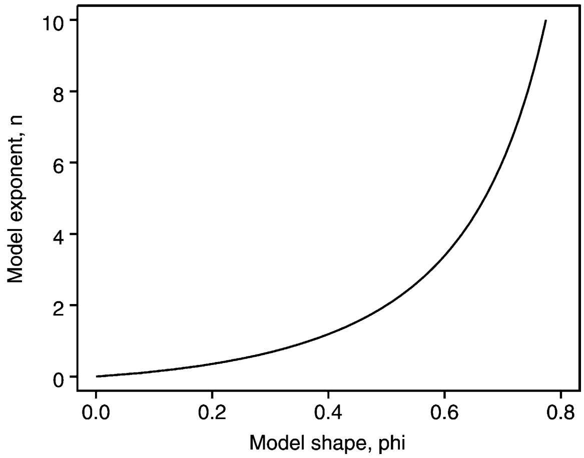

Fig. 2. Relationship between exponent n in generalized production model and model shape expressed as $\phi \equiv B_{\mathrm{MSY}}/K$. Singularity at $n = 1$ has been removed.

2.2. Fitting methods

The fitting method used for both models was that of Pella (1967) as revised by Prager (1994), also described by Quinn and Deriso (1999), and implemented as a generalized version of the computer program ASPIC (Prager, 1995). The method includes estimation conditional on observed catch, assuming lognormal observation error in fishing effort rate f_t in year t. The objective function was least-squares, with residuals computed as $r_t = \log(f_t) - \log(\hat{f}_t)$; under the assumptions used, this provides maximum-likelihood estimates. When conditioning on catch, such residuals in f are equivalent to negative residuals in the abundance index; thus, a residual here has negative sign when the observed abundance index is higher than the corresponding modeled value.

Continuous-time versions of both models were used. The logistic model used the analytical formulation of Prager (1994); the generalized model used a numerical approximation of the catch equation with between 24 and 100 steps per year, depending on estimated rate of change in population size. To condition on catch, the model biomass was projected forward each year and a corresponding trial value of F_t adjusted until predicted and observed catches matched (within some small tolerance), at which point the trial value was accepted as the model estimate. Parameter estimates from the logistic model were used as starting guesses in fitting the generalized model.

All nonlinear optimization algorithms tend to stop at local minima. Here, the polytope algorithm (Nelder and Mead, 1965; Wright, 1996) was used for minimization, and to increase the chance of finding a global minimum, parameter estimates were constrained to realistic values and the algorithm was repeatedly restarted after a minimum was located.

Monte Carlo searches of the parameter space were interspersed between restarts in a further attempt to avoid local minima. Three successive restarts to essentially the same estimates—five restarts when using least median of squares (described below)—were required before a set of estimates was accepted.

2.3. Bootstrap

Nonparametric bootstrapping (Efron, 1982; Efron and Gong, 1982; Stine, 1990) was used to estimate confidence intervals on quantities of interest by the BC (bias-corrected percentile) method of Efron (1982). The major advantages of the bootstrap in this application are its flexibility and relative freedom from distributional assumptions. Usually, bootstrapping is performed by resampling data observations; however, in fitting time series, the order of data must be retained. To preserve ordering, this study used a bootstrap method that combines the ordered predicted values from the original fit with residuals chosen at random from that fit (Efron and Tibshirani, 1986). Such bootstrap results are conditional on validity of the original fit.

2.4. Simulation

A simulation study was conducted to further compare accuracy and precision of the two estimators in a management context. The advantage of such a study is that estimates can be compared directly to the simulated ‘‘truth’’.

Simulations were based on the estimated stock trajectory from a trimmed generalized fit, described in Section 3.3. Simulated data sets were constructed using recorded data on catch, but replacing observed abundance-index values with estimates from that fit. Missing abundance-index values in the real data were treated as missing in the simulations. In each data set, simulated lognormal observation error was added to the abundance index. A total of 600 such data sets were generated, including 200 series of observation error at each of three coefficients of variation (CVs): 2, 10, and 25%. The first of these represents an almost ideal case, unattainable in an actual assessment. The final case was a guess at the magnitude of variability in real data. The remaining case was used to provide an intermediate value that might help reveal patterns in results.

Logistic and generalized models were fit to each simulated data set, and estimates of management-related quantities were recorded for each trial; i.e., estimates of MSY; relative fishing mortality rate in the final year of data, $F./F_{\mathrm{MSY}}$; and relative stock biomass after the final year, $B./B_{\mathrm{MSY}}$. Estimates of model shape ϕ from the generalized fits were also recorded.

Choice between models is often based on a statistical test. To examine the usefulness of that procedure in production-model choice, an F-ratio test was conducted comparing logistic and generalized fits of each simulated data set (under $H_0\colon \phi = 0.5$) and significance probability of each test was recorded.

Analysis of simulation results focused on estimation accuracy and precision, presented both through box–whisker plots for all combinations of CV and model shape and through statistics on simulations with 25% observation error, thought to be the most realistic of the levels simulated. Two statistics based on medians were computed, as more resistant to outliers than the usual statistics based on means.

To express accuracy of estimates, I computed median proportional error (MPE). For estimated quantity x with true value x^* and a set of estimates $\hat{x}_n, n = \{1, \ldots, N\}$, MPE was computed

$$\mathrm{MPE}(x) = \mathrm{median}_{n=1}^{N} \frac{\hat{x}_n - x^*}{x^*}. \tag{5}$$

Thus MPE is a resistant statistic similar to proportional bias, and one might expect an estimator with very low MPE to produce estimates above the true value about half the time and below it half the time.

Root-mean-squared error (RMSE; Kotz and Johnson, 1988) is a widely used measure of estimation merit that accounts for both accuracy and precision. For a set of N estimates of x as above, it is computed $\mathrm{RMSE}(x) = \left(\sum_{n=1}^{N}(\hat{x}_n - x^*)^2/N\right)^{1/2}$. An analogous measure, but more resistant and expressed scaled to the true value, is median proportional unsigned error (MPUE), defined here as

$$\mathrm{MPUE}(x) = \mathrm{median}_{n=1}^{N} \left| \frac{\hat{x}_n - x^*}{x^*} \right|. \tag{6}$$

The interpretation is that, on similar data, estimation error without respect to sign should be worse than MPUE about 50% of the time.

3. Application to data on swordfish

3.1. Nonlinear least-squares fitting

Initially, both models were fit to the ICCAT data set (Table 1) using the estimators described earlier with standard nonlinear least-squares (NLLS) as the objective function. Initial stock biomass was fixed at 0.875 *K* (ICCAT, 2000), a value believed to approximate the lightly exploited stock of 1950. Fixing the initial biomass in this way can increase estimation precision, possibly at the cost of some bias (Punt, 1990). Standard NLLS gave similar estimates of MSY, but widely different estimates of stock status, between models (Fig. 3a and b; Table 2), with estimates from the generalized model being more optimistic than estimates from the logistic model and also more optimistic than recent assessments of the stock (ICCAT, 2000). Two troubling aspects of the initial results were the large differences in estimates between models and the poor precision of estimates from each model (Table 2).

Further examination of results, prompted by the preceding issues, revealed several large residuals, the largest of which was associated with an extreme abundance-index value at the start of the series in 1963 (Fig. 3c and d). That value is questionable on purely biological grounds: If the value accurately represents relative abundance, the stock declined by about 65% from 1963 to 1964 and 77.5% from 1963 to 1965. Given the general life history of swordfish, which is characterized by only moderately variable recruitment and well-developed age structure even under exploitation, and the moderate level of fishing at the time, such a decline seems unlikely. The alternative is that the abundance index in 1963 did not represent true relative abundance. That could be the case, even if the index was correctly based on catch per effort in the fishery, if catchability in 1963 was higher than in following years, perhaps due to an unknown factor related to initiation of the longline fishery.

3.2. Outlier detection via resistant fitting

Given the concerns listed above, and because the 1963 abundance-index value seemed questionable at best, further analyses were done, starting with application of an objective method for identifying statistical outliers. The strategy described by Rousseeuw and Leroy (1987) for outlier detection comprises (1) use of a resistant fitting method to increase residuals of points not fitting the data–model combination well, (2) labeling as outliers points with the most extreme residuals, coupled with examination of outliers on substantive grounds, and (3) refitting the model with such outliers removed. Applications of this sequence in fisheries were given by Restrepo and Powers (1997).

The resistant fitting method used here was least median of squares (LMSs), which is resistant to 50% contamination of data by outliers. The method, described in detail by Rousseeuw and Leroy (1987), differs from standard NLLS only in that the quantity minimized is the median of squared residuals rather than the sum of squared residuals. The method has been advocated in fisheries problems by Chen and Paloheimo (1994); a similar method was used by Restrepo and Powers (1997).

Following application of LMS, all residuals were scaled by the two-stage resistant procedure of Rousseeuw and Leroy (1987; Chapter 5). In the first stage of that procedure, an initial scale estimate s^0 is computed from the minimal median residual and a finite-sample correction factor:

$$s^0 = 1.4826\left(\frac{1+5}{N-p}\right)\sqrt{\tilde{r}^2},$$

where N is the sample size, p the number of parameters estimated, and $\tilde{r}^2$ the median squared residual from the LMS fit. In computing the final scale estimate s^*, only the N^* residuals are used for which $|r/s^0| \le \xi$, where ξ is determined by the desired tolerance for outlying data. The final scale estimator computed from such residuals is

$$s^* = \sqrt{\frac{\sum^{N^*} r^2}{N^* - p}},$$

which is essentially a trimmed estimate of model standard error.

Given the final scale estimator s^* and the value of ξ set by the investigator, outliers are defined as points with $|r/s^*| > \xi$. Rousseeuw and Leroy (1987) suggest $\xi = 2.5$. However, fishery data are inherently more variable than most experimental data, so the value

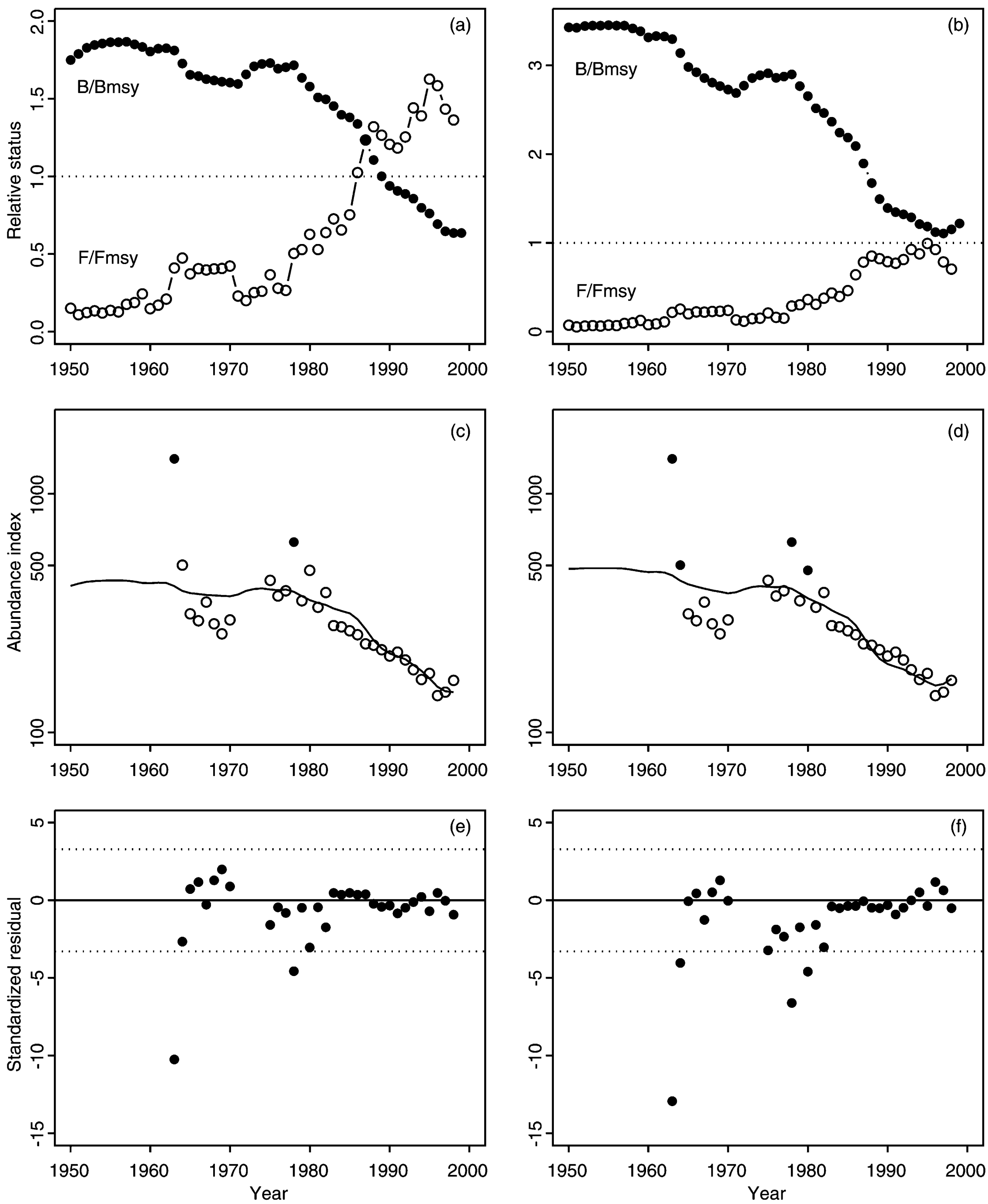

Fig. 3. Initial production model results for swordfish in north Atlantic Ocean. Panels (a) and (b), trajectories of relative biomass and fishing mortality rate estimated from logistic and generalized models, respectively, by NLLS; (c) and (d), observed (circles) and estimated (lines) abundance index for corresponding model fits, with filled symbols marking values designated as outliers in subsequent resistant (LMS) fits of same models; (e) and (f), standardized residual plots from resistant fits of same models, with dotted lines dividing outliers from other values.

Table 2
Goodness-of-fit of and estimates from production models of swordfish in north Atlantic Ocean[a]

Model shape	Nonlinear least-squares		Trimmed least-squares	
	Logistic[b]	Generalized	Logistic[b]	Generalized
R^2 in abundance index (%)	28.9	33.9	72.6	82.3
MSY	14,070 7320–16,160	14,510 5924–15,800	13,750 8940–15,310	14,040 10,370–15,940
$\phi \equiv B_{\mathrm{MSY}}/K$	0.500 –	0.256 0.153–0.287	0.500 –	0.545 0.336–0.627
$B_{1999}/B_{\mathrm{MSY}}$	0.635 0.503–0.999	1.22 0.966–2.33	0.719 0.622–0.882	0.709 0.589–0.921
$F_{1998}/F_{\mathrm{MSY}}$	1.36 0.784–2.08	0.707 0.292–0.902	1.24 0.962–1.82	1.23 1.03–1.72
Y_{1999} at F_{MSY}	8937 5566–16,270	17,690 13,960–39,440	9880 6576–13,230	9956 6934–12,120

[a] Trimmed fits are with 2 years' data dropped (logistic model) or 4 years' data dropped (generalized) as statistical outliers. Bootstrapped 80% confidence intervals given below estimates. Values of R^2 reflect differing sample sizes among models due to outlier removal, and are thus not strictly comparable.

[b] Model shape parameter $\phi = 0.5$ by definition for logistic model.

$\xi = 3.29$ was adopted here; that value defines a two-sided tail probability of 0.001 in a standard normal distribution.

Statistical outliers were identified in both models in accordance with the preceding procedure and LMS fitting. In the logistic model, two outliers were identified: data for 1963 and 1978, with scaled residuals (r/s^*) of -10.26 and -4.56, respectively (Fig. 3c and e). In the generalized model, four outliers were identified: 1963, 1964, 1978, and 1980, with scaled residuals of -12.93, -4.03, -6.59, and -4.59, respectively (Fig. 3d and f).

3.3. Trimmed estimates

Both models were refit after eliminating the outliers identified above. In doing this, outlying values of the abundance index were treated as missing data, but corresponding data on catch were retained. Thus, slightly different sets of data were used in fitting the two models, the logistic model admitting 2 more years of relative-abundance data than the generalized model. Following Restrepo and Powers (1997), I refer to the final fitting procedure as trimmed nonlinear least-squares, or TNLLS, and to estimates from this procedure as trimmed estimates.

The TNLLS procedure provided much better fit between model and data, as would be expected when outliers are removed. Estimated precision of corresponding trimmed estimates was also better (Table 2, Fig. 4).

Trimmed logistic estimates were similar to the original logistic estimates, although the trimmed estimates were slightly more optimistic about stock status and considerably more precise. Trimmed generalized estimates were quite different from the original NLLS generalized estimates, but were closer to both sets of logistic estimates. Estimates of MSY were less affected by removal of outliers than estimates of other quantities (Table 2).

Both logistic and generalized trimmed estimates describe the stock as somewhat depleted, with B_{1999} at about 70% of B_{MSY} and F_{1998} at about 125% of F_{MSY}. Measures of stock status are expressed relative to benchmarks to increase precision by cancellation of estimated catchability $\hat{q}$ and because the ratios are of interest to management (Prager, 1994). Both models estimate that if the stock were fished at F_{MSY} in 1999, the yield would be about 9900 t, considerably less than recent catches. A constant-effort policy based in F_{MSY} is the upper limiting case of recent US technical guidelines (Restrepo et al., 1998), and is used here

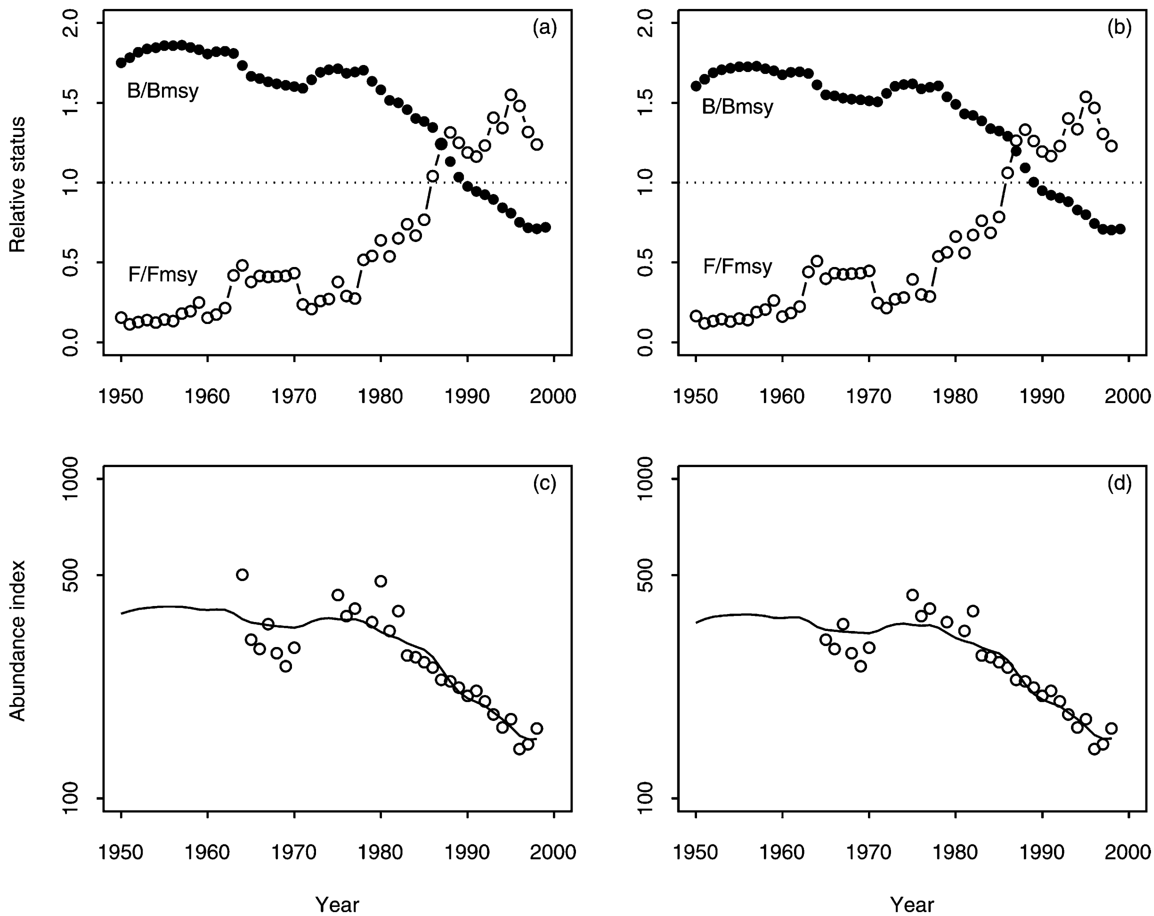

Fig. 4. Final production model results for swordfish in north Atlantic Ocean. Panels (a) and (b), trajectories of relative biomass and fishing mortality rate estimated from logistic and generalized models, respectively, by trimmed nonlinear least-squares; (c) and (d), corresponding observed (circles) and estimated (lines) abundance indices.

to represent the class of constant-effort harvesting policies.

4. Simulation results

Results of the simulation are presented both graphically (for all values of simulated observation error) and in tabular form (for 25% CV of observation error only), to provide complementary views of estimation performance. The graphical presentation (Fig. 5) gives estimation errors in the original units, while tabulated statistics (Table 3) are expressed as percentages of true values. Because all 600 data sets had the same true values of management quantities, patterns of errors are independent of that difference in scaling.

Not all 600 simulated data sets are included in these results, as it was not possible to obtain estimates from every data set. At 2% error, all fits were successful; at 10% error, all logistic fits were successful, but two generalized fits hit bounds on $\hat{\phi}$; at 25% error, there were 11 failures to reach convergence on the logistic model and 50 failures on the generalized model (11 trials abandoned when the simpler logistic fit failed and 39 bounds hits on $\hat{\phi}$). Thus, it was not possible to fit the generalized model in 25% of cases with the highest level of observation error, and it was not possible to fit the logistic in about 5% of such cases.

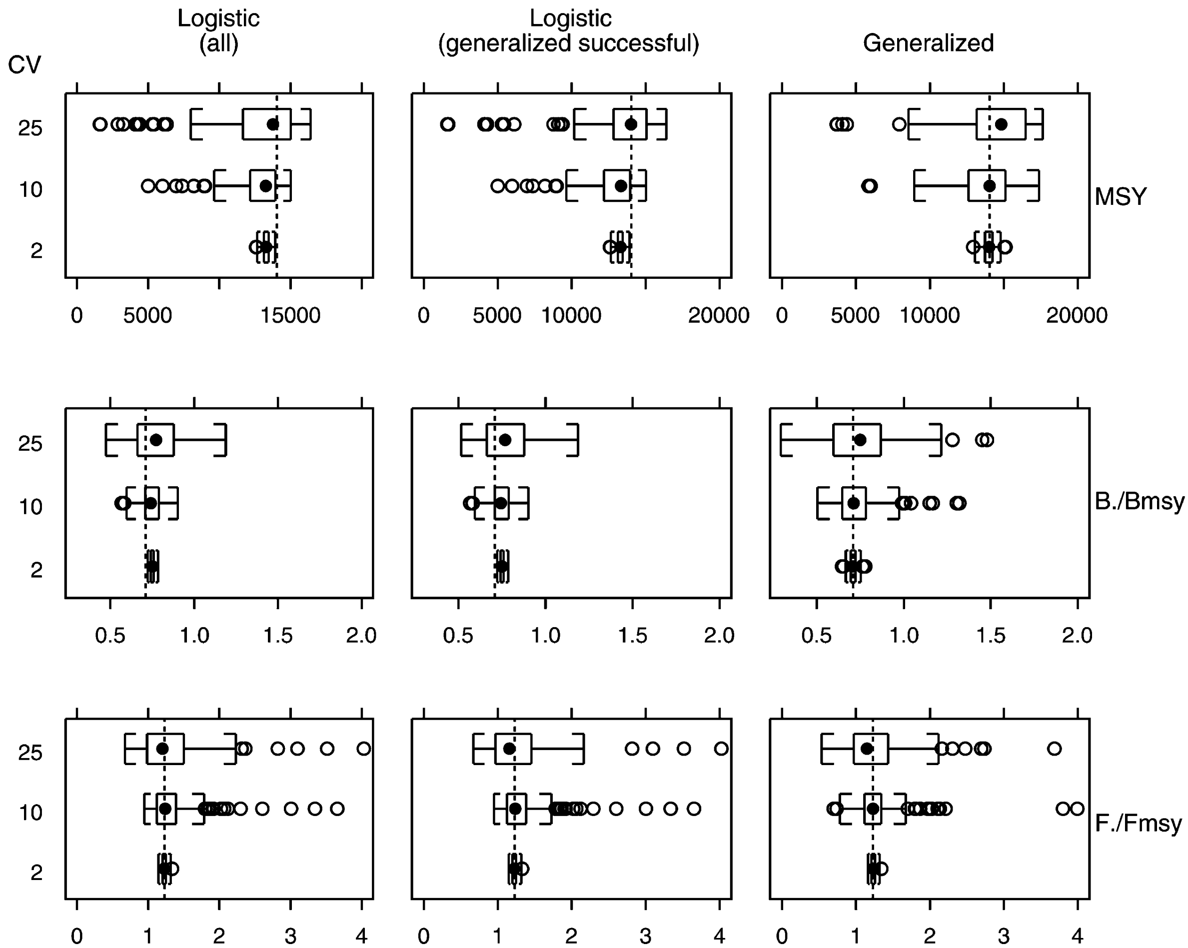

Fig. 5. Distributions of logistic (left; center) and generalized (right) production model estimates from analyses of simulated data. Simulated data were based on trimmed generalized production-model fit of real data on swordfish in north Atlantic Ocean, following outlier detection and removal. Simulated observation error added at three levels, CV $\in \{2, 10, 25\%\}$. Dashed lines are "true" values used in simulation. Box–whisker plots depict median (filled circle), interquartile range (box), general data range (whiskers), and extreme observations (open circles).

Table 3
Performance of logistic and generalized production-model estimators applied to simulated data resembling swordfish in north Atlantic Ocean[a]

Estimated quantity	MPE (%)		MPUE (%)	
	Logistic (1)	Generalized (2)	Logistic (3)	Generalized (4)
MSY	0.0	5.8	7.8	13.5
$B_{1999}/B_{\mathrm{MSY}}$	8.4	5.8	14.1	19.5
$F_{1998}/F_{\mathrm{MSY}}$	−5.5	−6.6	19.9	20.0

[a] Results shown for 150 simulations with 25% observation error in which generalized estimator converged on a solution. MPE is a statistically resistant analog of proportional bias; MPUE, a resistant analog of proportional RMSE (see Section 2.4 for definitions).

Median estimation error in MSY was comparable between estimators at all levels of observation error (heavy black dots in top three panels of Fig. 5), with the logistic estimator performing slightly better on data with 25% CV (Table 3), and the generalized estimator performing slightly better at lower levels of observation error. When considering only the subset of results in which both estimators were successful estimates of MSY from the logistic estimator were considerably more precise, as judged by extent of the central boxes and whiskers in Fig. 5. Even when all successful logistic estimates are compared to the lower number of successful generalized estimates, the logistic estimator appears more precise.

Of the three management quantities considered, increased estimation precision from the logistic estimator was most pronounced in $B./B_{MSY}$. Logistic estimates of this quantity were slightly less accurate, but considerably more precise, than generalized estimates, particularly at the 25% level of observation error (Fig. 5, Table 3).

Median estimation error in $F./F_{MSY}$ was comparable from the two estimators. Logistic estimates of $F./F_{MSY}$ were slightly more precise than corresponding generalized estimates (Fig. 5), especially at 25% observation error (Table 3). Estimation of this quantity appears relatively insensitive to the choice of estimator.

Estimation of model shape ϕ was imprecise at the 10 and 25% levels of observation error, with no estimates obtained at $\hat{\phi} < 0.25$ or $\hat{\phi} > 0.75$ because of bounds set on $\hat{\phi}$ to facilitate estimation (Fig. 6). In some noisy data sets, model fit is optimized by increasing or decreasing $\hat{\phi}$ without limit; estimates at either bound were thus discarded as invalid. This is not meant to imply that in some species the true value of ϕ might not fall outside these limits.

Estimates of ϕ were correlated with estimates of other quantities (Fig. 7). Estimates of MSY tended to increase with increasing $\hat{\phi}$, as did estimates of $F./F_{MSY}$. Estimates of $B./B_{MSY}$ were strongly negatively correlated with $\hat{\phi}$ and the strength of that correlation may explain the relatively high dispersion noted above in estimates of $B./B_{MSY}$.

MPUE was computed as a robust measure of estimation merit reflecting both accuracy and precision. Results favor the logistic model, which produced lower MPUE in MSY and $B./B_{MSY}$; the two models produced essentially the same MPUE in $F./F_{MSY}$ (Table 3).

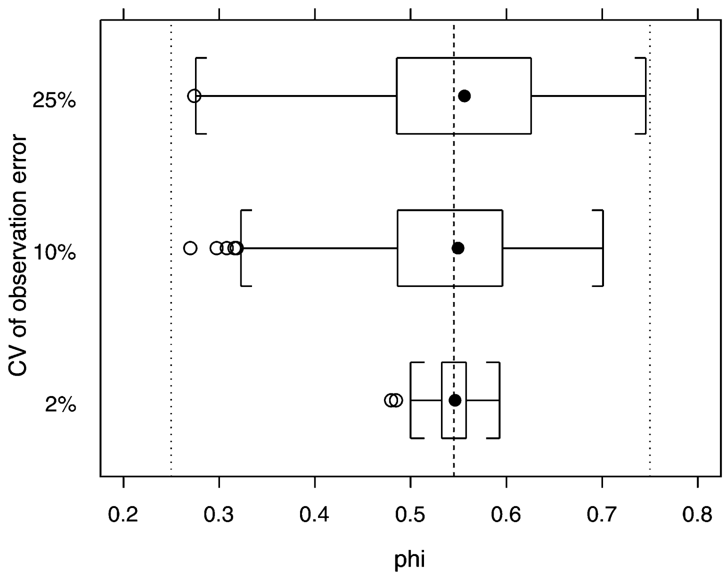

Fig. 6. Distribution of estimates of model shape $\phi \equiv B_{MSY}/K$ from 600 simulated data sets patterned on swordfish in north Atlantic ocean; shown by CV of simulated lognormal observation error. Dashed line is "true" value of ϕ used in simulation; dotted lines are constraints placed on $\hat{\phi}$ to facilitate estimation. Box–whisker plots depict median (filled circle), interquartile range (box), general data range (whiskers), and extreme observations (open circles).

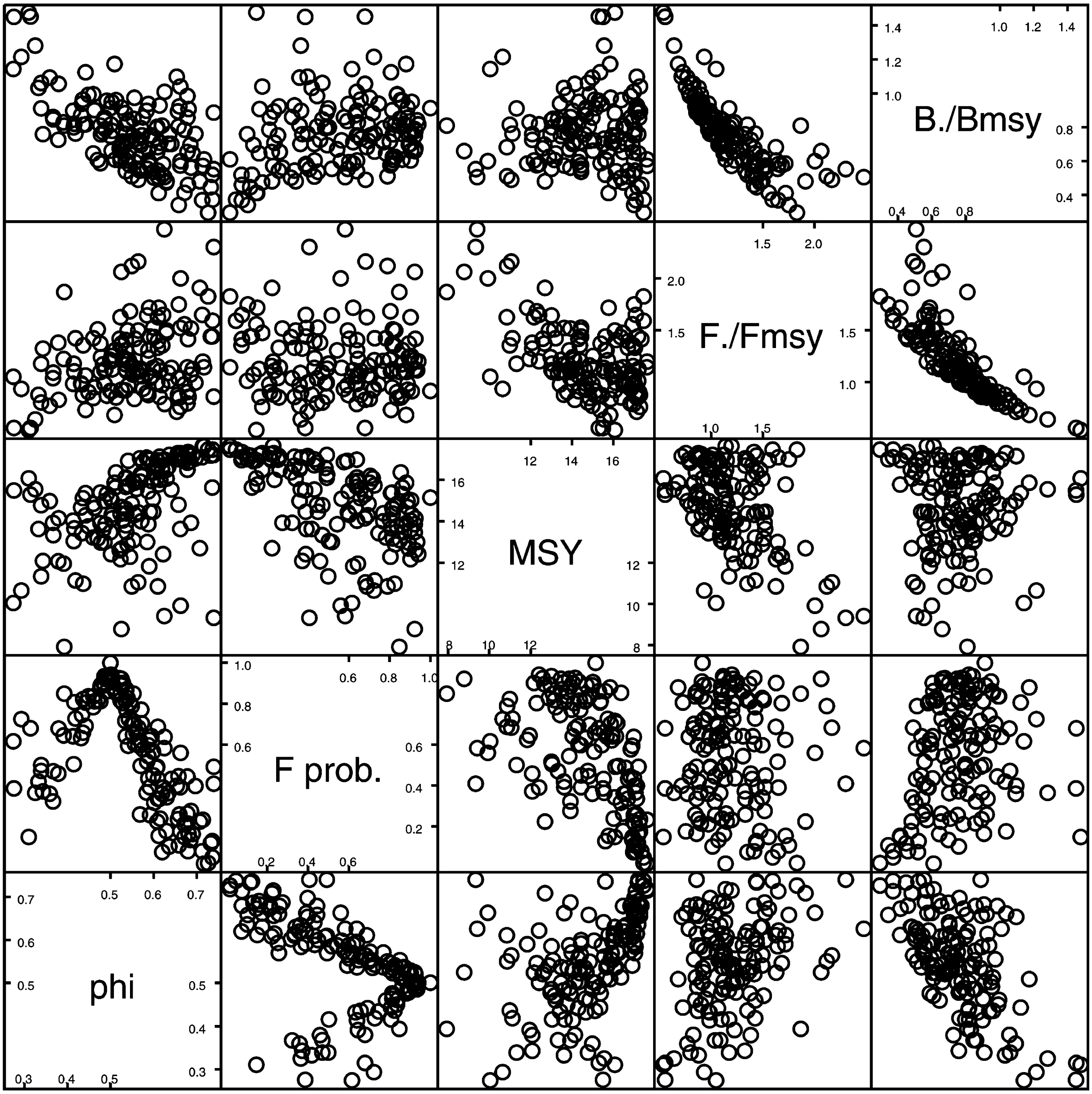

Fig. 7. Correlations among estimated quantities in simulations based on swordfish in north Atlantic Ocean. Estimates are from generalized production model fits of 150 data sets based on one underlying population trajectory with different realizations of simulated lognormal observation error at 25% CV. Five cases with estimated $F./F_{\mathrm{MSY}} > 2.5$ are off scale; MSY divided by 1000 for plotting. "F-prob" is significance probability of an F-ratio test of H_0: $\phi = 0.5$.

4.1. Use of statistical tests for model choice

The significance probability (P) of an F-ratio test was not particularly informative for choosing model results with better management estimates. The lowest estimated P values consistently occurred when $\hat{\phi}$ was furthest from the null value of 0.5 (Fig. 7). With true model shape $\phi = 0.55$ in the simulated populations,

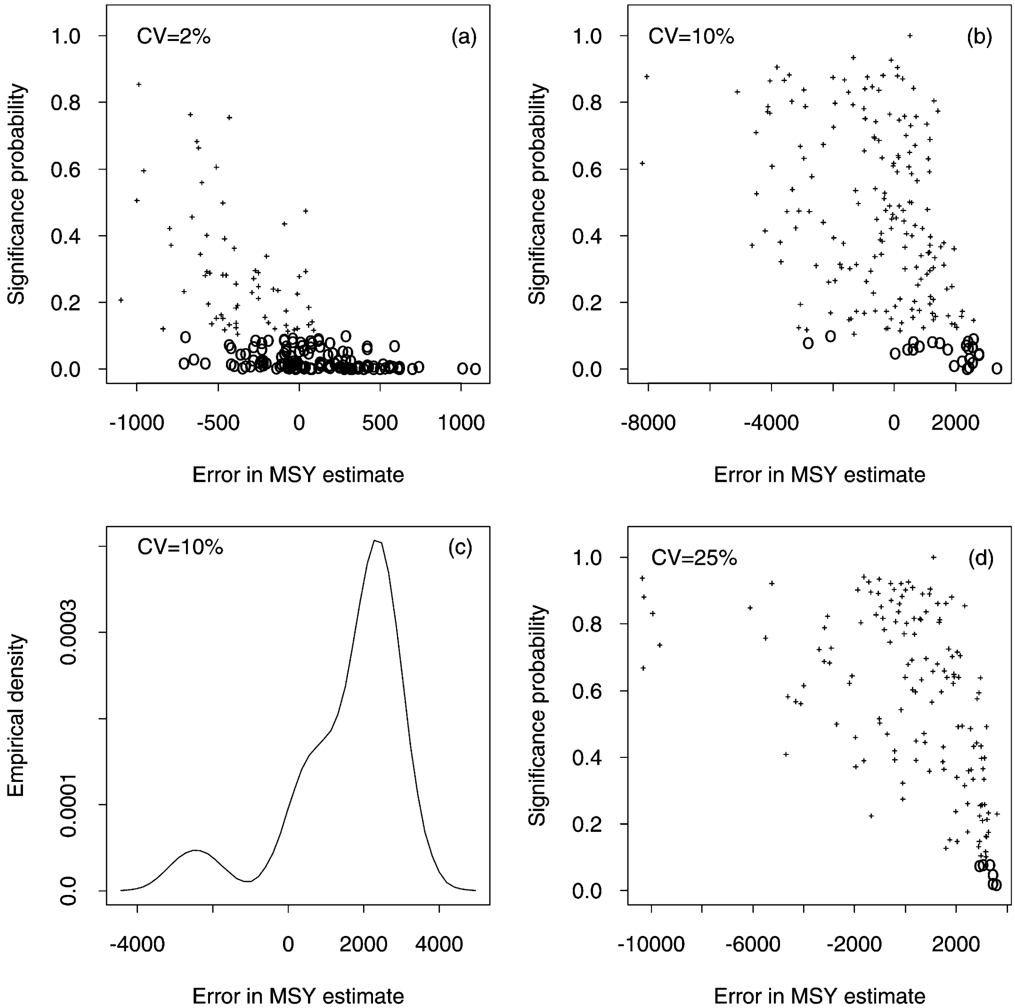

Fig. 8. Use of F-ratio test to choose between logistic production model ($\phi \equiv 0.5$) and generalized model (ϕ estimated). Significance probabilities P for tests of H_0: $\phi = 0.5$ on simulated data with true $\phi = 0.55$. Panels (a), (b), and (d): relationship of P to error in estimate of MSY from data with simulated observation error at three levels. Small crosses mark cases not significant at $P < 0.1$, larger circles mark significant cases. Panel (c): empirical probability density of error in MSY estimates for significant cases in panel (b).

the most statistically significant results (smallest values of P) were estimates of management quantities at extreme values (Fig. 7).

Significance tests were conducted at $\alpha = 0.1$ for the 600 simulations. As the CV of observation error in the simulated data was increased to levels believed realistic, the test became less useful. At 2% error, the test tended to choose model estimates approximately centered around the correct value of MSY (Fig. 8a). However, at 10% observation error, significant test results were most common at estimates of MSY about 2000 t yr^{-1} higher than the true (simulated) value of about 14,000 t yr^{-1} (Fig. 8b and c). The few significant test results at 25% observation error were located even further from the correct value of MSY (Fig. 8d).

5. Discussion

The goal of this methodological study was to examine relative strengths and weaknesses of logistic and generalized production models in assessment of this stock. A secondary goal was exploring the utility of robust regression in treatment of apparent outliers in the data. (Although this study also resulted in estimates of stock status, the author does not consider them to supersede the usual assessments of this stock, which include varied analyses and extensive synthesis by an international panel of specialists.) Several insights have been gained in this investigation.

5.1. Longline-based abundance indices

As noted by Au (1986), abundance indices derived from longline fishery data tend to show a marked decline in the first few years of fishing, a decline that far exceeds the expected effects of fishing. In the longline-based index for north Atlantic swordfish, a decline of 77.5% occurs between 1963 and 1965, when catches were below estimated MSY levels (Tables 1 and 2) and estimates of fishing mortality rate were below F_{MSY} (panels a and b of Figs. 3 and 4). Although some reduction of the stock certainly occurred, it seems unlikely that the magnitude of the index decline was matched by a proportional decline in stock biomass. A general alternative to proportionality is that catchability in the first few years of a longline fishery tends to be unusually high. To my knowledge, a definitive explanation of that phenomenon has not been developed; a hypothesis due to J. Radovich is that initial crowding may cause an increased tendency to take the hook (A.D. MacCall, pers. comm.). Because fishery management often attempts to restore depleted stocks to some fraction of their initial condition, early data from fisheries are particularly critical, yet we do not always understand the meaning of initial observations in such time series. Here, a high initial value was rejected as a statistical outlier, an approach that highlights the problem but does not overcome it. A valid quantitative model of the elevated catch per effort observed in the first few years of longline fisheries would constitute an important theoretical advance in assessment and management of stocks taken on longlines. Unfortunately, as more and more time elapses since the observation of those conditions, it seems increasingly difficult to study the problem, at least in a way that might allow field verification.

5.2. Resistant fitting methods

Resistant fitting methods were found useful here to identify outliers. Because it was not possible to correct or adjust the outliers, they were eliminated from the analysis. Whether resulting estimates were more or less biased than those including the outliers cannot be determined, and would seem to depend on the unknown mechanism generating the outliers. In general, it seems preferable to use resistant fitting to identify and correct outliers if possible rather than simply to discount them (as would be done if the resistant fit itself were used for estimation) or eliminate them (as was done here). It also seems that resistant methods would best be used in fishery management in conjunction with some form of sensitivity analysis; i.e., fits with and without suspected outliers will ideally be examined as to precision and bias of results. Here, resistant methods gave appreciably more precise estimates; it seems worth investigating whether such methods might also reduce the variation in estimates over time (from one assessment to the next).

5.3. Shape of the production curve

The true shape of the production curve for this stock is unknown. Although the shape was estimated, its sensitivity to observation error in the data (Fig. 6) implies that the estimate is provisional at best. At least two authors have attempted to estimate ϕ for swordfish from demographic considerations in accordance with theory advanced by Fowler (1981, 1988). Garcia-Saez (1997) applied age- and stage-structured matrix models with harvesting and arrived at the range $0.68 \leq \hat{\phi} \leq 0.97$ when considering total stock biomass, and $0.59 \leq \hat{\phi} \leq 0.65$ when considering reproductive (mature) biomass. The author concluded that "Growth rates calculated here imply that not more than 40% of the population could sustainably be harvested (inflection point of $0.6\,K$)...". In a Bayesian analysis based on a modified generalized production model, McAllister et al. (2000) established a joint prior distribution on the intrinsic rate of increase r and the shape expressed as an exponent, as in Eq. (2). That prior distribution was developed from the empirical relationship of Fowler (1988); the prior median estimate was $\hat{\phi} = 0.43$, and the posterior distribution was centered at $\hat{\phi} = 0.4$.

Neither of the preceding estimates is conclusive. Each is grounded in quantitative theory but necessarily involves many assumptions. For example, specific forms of stock–recruitment relationship were assumed, and specific forms of other demographic functions were used, generally based on studies at a limited range of population sizes. In this study, the TNLLS 80% confidence interval on $\hat{\phi}$ (Table 2) encompasses both the central estimates of McAllister et al. (2000) and most of the range of estimates of Garcia-Saez (1997) based on mature biomass, but does not include the latter's higher estimates based on total biomass.

The assumption $\phi = 0.5$, which results in the logistic model, was used by Lotka (1924) not for biological realism but as a central approximation possessing mathematical and conceptual simplicity. Thompson (1992) found through algebraic manipulation that under simple dynamic pool models with nondecreasing stock–recruitment relationships (such as the Beverton and Holt (1957) model), it is always true that $\phi < 0.5$. That seems to set an upper limit on ϕ for stocks meeting the specified assumptions, but a lower limit can also be proposed based on properties of the generalized model itself. In the generalized model with $n \leq 1$, equivalent to $\phi \lessapprox 0.37$, a positive (and substantial) sustainable yield is available even as F approaches infinity, because the stock's rate of increase becomes infinite as the stock size goes to zero. Clearly, that property is not exhibited by real fish stocks, which often exhibit reduced productivity at low stock levels. To eliminate that model property, McAllister et al. (2000) advocated a hybrid of the Fletcher (1978) formulation and the logistic model, an approach that provides flexible model shape while avoiding an infinite rate of increase. Another approach would be to impose on the generalized model (Eq. (2)) the constraint $\phi > 0.37$.

Thus, two closely related issues exist about estimating the shape of the production curve externally (to a production model), whether the goal is point estimation of ϕ or estimation of a prior distribution for a Bayesian approach. The first issue is determining an estimation procedure for a given population, i.e., choosing the models and assumptions to be used. For swordfish, broader study of that problem is required, including study of sensitivity to changes in assumptions. The second issue is that all generalizations about allowable or probable values of ϕ are rooted in specific models. The degree to which those models are representative of real fish stocks for this purpose is unknown.

5.4. Choices for assessment work

What can be said about appropriate choice of estimators when the goal is conducting an assessment for management? Some data sets are insufficiently informative to estimate ϕ and, as has been seen, when ϕ can be estimated, it may be difficult to have confidence in the results. When evaluated on the swordfish data (real and simulated), the generalized model with estimated exponent was quite sensitive to outliers; the logistic model was much less sensitive. For that reason, the logistic production model can be recommended as a central approximation that will likely provide more precise and stable estimates, or provide estimates when none can be obtained from the generalized model. In addition, the logistic model may be superior (as here) to the generalized model in terms of MPUE, a measure considering both accuracy and precision.

Given the range of previous estimates of ϕ for swordfish in the north Atlantic Ocean, I believe the simulation results in this paper are applicable to that stock. The present results, obtained on a production curve skewed slightly to the left, suggest that the logistic may also outperform the generalized estimator on production curves in the neighborhood of the logistic, including curves skewed slightly to the right. It would require more extensive simulations to establish that supposition conclusively and to establish whether the superior performance of the logistic estimator applies across the range of plausible shapes of production curve for all species.

In general, it seems likely that better estimates could be obtained from the generalized model by specifying a priori the correct value of ϕ. The difficulty, of course, is knowing the correct value of ϕ. As noted earlier, more research on that topic is needed, both on theory and specifically on swordfish.

As noted above, the F-ratio test may not be useful at choosing the model form with better management estimates; here, it tended to choose the generalized model only when $\hat{\phi}$ was furthest from the null value of 0.5. I would expect other test statistics based on goodness of fit (e.g., Akaike, 1974; Schwarz, 1978) to behave similarly: when comparing logistic and generalized models, large increases in goodness of fit will occur most often when the generalized estimate of ϕ is most different from the logistic assumption $\phi = 0.5$. Because, as noted earlier, estimates of ϕ can be quite imprecise and sensitive to observation error (Fig. 6), statistical tests may prove of very limited utility in choosing between these two models. The simulation study here was a relatively difficult test of utility, as the true model shape was close to the logistic, and significance tests may perform better when the true shape is farther from the logistic. Such tests are used only when the true shape is unknown, however, so that consideration may ultimately be unimportant. Until more extensive studies have been made of the utility of significance tests in this context, I recommend that they be considered unreliable.

The difficulty in fitting the generalized model (with estimated shape) has often been cast as a technical problem of overly correlated parameters or difficulty in obtaining convergence from a nonlinear optimization algorithm. Even when the data do sufficiently define the shape of the production curve, technical difficulties in estimating that shape can be substantial. Nonetheless, this study has led to a different conclusion, that the underlying problem with the generalized model is structural: the model is extremely sensitive to outliers. Fishery data sets tend to be short, noisy, time series; estimates of ϕ are quite sensitive to noise; estimates of management quantities propagate that sensitivity. Thus, the problem with the generalized model is more fundamental than it is technical, it is intertwined with the noisy nature of fishery data, and the warning of Hilborn and Walters (1992) cited in Section 1 seems quite correct.

In the simulation study, estimates from the generalized model were found of comparable accuracy to those from the logistic model, but the generalized model was quite sensitive to outlying observations and characterized by higher MPUE. Because wider dispersion in estimators can lead to misleading parameter estimates in practice, I conclude that if the generalized production model with estimated shape is used in swordfish management, it should be used with skepticism and only in conjunction with the more robust logistic form. It also seems useful to establish formal procedures for analysis of statistical outliers and to examine sensitivity of estimates to assumptions made about model shape.

Acknowledgements

I thank D. Ahrenholz, E. Bjorkstedt, D. Colby, P. Mararro, J. Powers, V. Restrepo, W. Richards, G. Scott, G. Thayer, D. Vaughan, E. Williams, and two anonymous referees for reviewing versions of the manuscript and for many useful comments. The assistance of E. Williams was invaluable in refining estimation software for the generalized model. I thank V. Restrepo of ICCAT for forwarding the data and for his thought-provoking comments on resistant modeling techniques. Remaining errors are of course the responsibility of the author. Two freely available software tools were used in the simulation: Recursive Replace (http://www.code66.com/tools/) and the Python scripting language (http://www.python.org/); I thank those whose labor has made those tools available. The work was supported by the US National Marine Fisheries Service, Southeast Fisheries Science Center, through the NOAA Center for Coastal Fisheries and Habitat Research.

References

Akaike, H., 1974. A new look at the statistical identification model. IEEE Trans. Auto. Contr. 19, 716–723.

Au, D., 1986. Interpretation of longline hook rates. Coll. Vol. Sci. Pap. ICCAT 25, 377–385.

Beverton, R.J.H., Holt, S.J., 1957. On the dynamics of exploited fish populations. Fish. Invest. Ser. 2, Mar. Fish. GB Minist. Agric. Fish. Food, Vol. 19, 533 pp.

Chen, Y., Paloheimo, J.E., 1994. Robust regression approach to estimating fish mortality rates with a cohort-based model. Trans. Am. Fish. Soc. 123, 508–518.

Efron, B., 1982. The Jackknife, the Bootstrap, and Other Resampling Plans. Soc. Industr. Appl. Math., Philadelphia, PA, 92 pp.

Efron, B., Gong, G., 1982. A leisurely look at the bootstrap, the jackknife, and cross-validation. Am. Statist. 37, 36–48.

Efron, B., Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statist. Sci. 1, 54–75.

Fletcher, R.I., 1978. On the restructuring of the Pella–Tomlinson system. Fish. Bull. 76, 515–521.

Fowler, C.W., 1981. Density dependence as related to life history strategy. Ecology 62, 602–610.

Fowler, C.W., 1988. Population dynamics as related to rate of increase per generation. Evol. Ecol. 2, 197–204.

Fox, W.W., 1970. An exponential surplus-yield model for optimizing exploited fish populations. Trans. Am. Fish. Soc. 99, 80–88.

Fox, W.W., 1975. Fitting the generalized stock production model by least-squares and equilibrium approximation. Fish. Bull. 73, 23–37.

Garcia-Saez, C., 1997. Assessing populations of swordfish from the north Atlantic: density dependence, maximum sustainable yield levels, and size-classified demographic models. Coll. Vol. Sci. Pap. ICCAT 44, 345–353.

Gavaris, S., 1988. An adaptive framework for the estimation of population size. Canadian Atlantic Fisheries Scientific Advisory Committee (CAFSAC), Res. Doc. 88/29, 12 pp.

Hilborn, R., 1990. Estimating the parameters of full age-structured models from catch and abundance data. Int. N. Pac. Fish. Comm. Bull. 50, 207–213.

Hilborn, R., Walters, C.J., 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics, and Uncertainty. Chapman and Hall, New York, 570 pp.

Hoey, J.J., Mejuto, J., Porter, J.M., Stone, H.H., Yokawa, K., 2000. An updated biomass index of abundance for north Atlantic swordfish 1963–1998. Coll. Vol. Sci. Pap. ICCAT 51, 1621–1636.

ICCAT (International Commission for the Conservation of Atlantic Tunas), 2000. Report for the biennial period, 1998–1999, Part 2 (1999), Vol. 2. Int. Comm. Cons. Atl. Tunas, Madrid.

Kotz, S., Johnson, N.L. (Eds.), 1988. Encyclopedia of Statistical Sciences, Vol. 8. Wiley, New York, 870 pp.

Lotka, A.J., 1924. Elements of Physical Biology. Reprinted 1956 as Elements of Mathematical Biology. Dover, New York, 465 pp.

Ludwig, D., Walters, C.J., Cooke, J., 1988. Comparison of two models and two estimation methods for catch and effort data. Nat. Res. Modeling 2, 457–498.

McAllister, M.K., Babcock, E.A., Pikitch, E.K., Prager, M.H., 2000. Application of a non-equilibrium generalized production model to south and north Atlantic swordfish: combining Bayesian and demographic methods for parameter estimation. Coll. Vol. Sci. Pap. ICCAT 51, 1523–1551.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Comp. J. 7, 308–313.

Pella, J.J., 1967. A study of methods to estimate the Schaefer model parameters with special reference to the yellowfin tuna fishery in the eastern tropical Pacific ocean. Dissertation. University of Washington, Seattle, WA.

Pella, J.J., Tomlinson, P.K., 1969. A generalized stock production model. Bull. Inter-Am. Trop. Tuna Comm. 13, 419–496.

Polacheck, T., Hilborn, R., 1987. A proposal to abandon the use of equilibrium methods for surplus production models. ICSEAF Document SAC/87/Doc/20, 11 pp.

Polacheck, T., Hilborn, R., Punt, A.E., 1993. Fitting surplus production models: comparing methods and measuring uncertainty. Can. J. Fish. Aquat. Sci. 50, 2597–2607.

Prager, M.H., 1994. A suite of extensions to a nonequilibrium surplus-production model. Fish. Bull. 92, 374–389.

Prager, M.H., 1995. User's Manual for ASPIC: A Stock-production Model Incorporating Covariates, Program Version 3.6x, 4th Edition. NMFS Southeast Fisheries Science Center, Miami Laboratory Document MIA-2/93-55. Available from the author.

Punt, A.E., 1990. Is $B_1 = K$ an appropriate assumption when applying an observation error production-model estimator to catch-effort data. S. Afr. J. Mar. Sci. 9, 249–259.

Punt, A.E., 1992. Selecting management methodologies for marine resources, with an illustration for southern African hake. S. Afr. J. Mar. Sci. 12, 943–958.

Punt, A.E., Hilborn, R., 1996. Biomass Dynamic Models: User's Manual. FAO Computerized Information Series (Fisheries), No. 10. FAO (Food and Agriculture Organization of the United Nations), Rome, 62 pp.

Quinn, T.J., Deriso, R.B., 1999. Quantitative Fish Dynamics. Oxford University Press, New York, 542 pp.

Restrepo, V.R., Powers, J.E., 1997. Application of high-breakdown robust regression to tuned stock assessment models. Fish. Bull. 95, 149–160.

Restrepo, V.R., Hoenig, J.M., Powers, J.E., Baird, J.W., Turner, S.C., 1992. A simple simulation approach to risk and cost analysis, with applications to swordfish and cod fisheries. Fish. Bull. 90, 736–748.

Restrepo, V.R., Thompson, G.G., Mace, P.M., Gabriel, W.L., Low, L.L., MacCall, A.D., Methot, R.D., Powers, J.E., Taylor, B.L., Wade, P.R., Witzig, J.F., 1998. Technical guidance on the use of precautionary approaches to implementing National Standard 1 of the Maguson–Stevens Fishery Conservation and Management Act. NOAA Technical Memorandum NMFS-F/SPO-31.

Rivard, D., Bledsoe, L.J., 1978. Parameter estimation for the Pella–Tomlinson stock production model under nonequilibrium conditions. Fish. Bull. 76, 523–534.

Roff, D.A., Fairbairn, D.J., 1980. An evaluation of Gulland's method for fitting the Schaefer model. Can. J. Fish. Aquat. Sci. 37, 1229–1235.

Rousseeuw, P.J., Leroy, A.M., 1987. Robust Regression and Outlier Detection. Wiley, New York, 329 pp.

Schaefer, M.B., 1954. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. Bull. Inter-Am. Trop. Tuna Comm. 1 (2), 27–56.

Schaefer, M.B., 1957. A study of the dynamics of the fishery for yellowfin tuna in the eastern tropical Pacific Ocean. Bull. Inter-Am. Trop. Tuna Comm. 2, 247–268.

Schnute, J., 1977. Improved estimates from the Schaefer production model: theoretical considerations. J. Fish. Res. Board Can. 34, 583–603.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6, 461–464.

Stine, R., 1990. An introduction to bootstrap methods: examples and ideas. In: Fox, J., Long, J.S. (Eds.), Modern Methods of Data Analysis. Sage, Newbury Park, CA, pp. 325–373.

Thompson, G.G., 1992. Management advice from a simple dynamic pool model. Fish. Bull. 90, 552–560.

Wright, M.H., 1996. Direct search methods: once scorned, now respectable. In: Griffiths, D.F., Watcon, G.A. (Eds.), Numerical Analysis 1995. Addison-Wesley/Longman, Harlow, UK, pp. 191–208.